Progress Report on AISR Grant NNG06GE71G
Principal Investigator: István Szapudi
Scalable Algorithms for Fast Analysis of Megapixel CMB Maps
and Large Astronomical Databases
Institute for Astronomy, Hawaii

Cumulative up to December 2009 (in Reverse Order)

We have introduced new extensions into our software pipeline based on SpICE that enables calculation of the cumulant correlator power spectrum. We demonstrated that it works in a real world environment by by calculating the 2-1 cumulant correlator power spectrum and WMAP3. A new method for calculating the full bispectrum based on our previous results on three-point functions is under way. We started to develop a new method, based on shrinkage estimators, which has a potential to improve statistical estimates in many areas of astrophysics. We first applied this idea for determining the covariance matrix of power spectrum measurement, and we are working on applying the same idea for C_l estimation of the CMB. We have been developing two new techniques to look for ISW effect in the CMB, one is searching for hot and cold spots, while the other is using a matched filter. We have developed Monte Carlo techniques for density reconstruction, as well as an algorithm quantifying higher order statistics using non-linear mappings.

1 Principal Results in 2009

1.1 Precise Density Estimation with a Monte Carlo Markov Chain Method

We developed a Monte Carlo algorithm to estimate the underlying dark matter density field using a galaxy catalog consisting of discrete objects possibly subject to photometric redshift errors. We published the results in Granett et al. (2009) that aimed at detecting a hypothesied large void towards the "Cold Spot" region in the CMB in the WMAP data. Our conclusion was that the present data excludes a supervoid on large redshift but further galaxy data will be needed to test whether there is a large void at law redshift.

1.2 A Non-linear Mapping of Higher Order Statistics into Second Order

In Neyrinck et al. (2009) we found that nonlinearities in the dark-matter power spectrum are dramatically smaller if the density field first undergoes a pair of nonlinear mappings: logarithmic and Gaussianization mapping. Such mapping results in algorithms that are as fast as two-point statistics, yet access to information hitherto only accessible to higher order methods in the (weakly) non-linear regime, thus it has great potential for application in large databases. We tested the idea using the Millennium simulation, finding that this procedure gives a power spectrum with a shape hardly departing from the linear power spectrum for $k \lesssim 1\ h^{-1}$Mpc at all redshifts. Also, this procedure unveils pristine Fisher information on a range of scales reaching a factor of 2-3 smaller than in the standard power spectrum, yielding 10 times more cumulative signal-to-noise at $z = 0$.

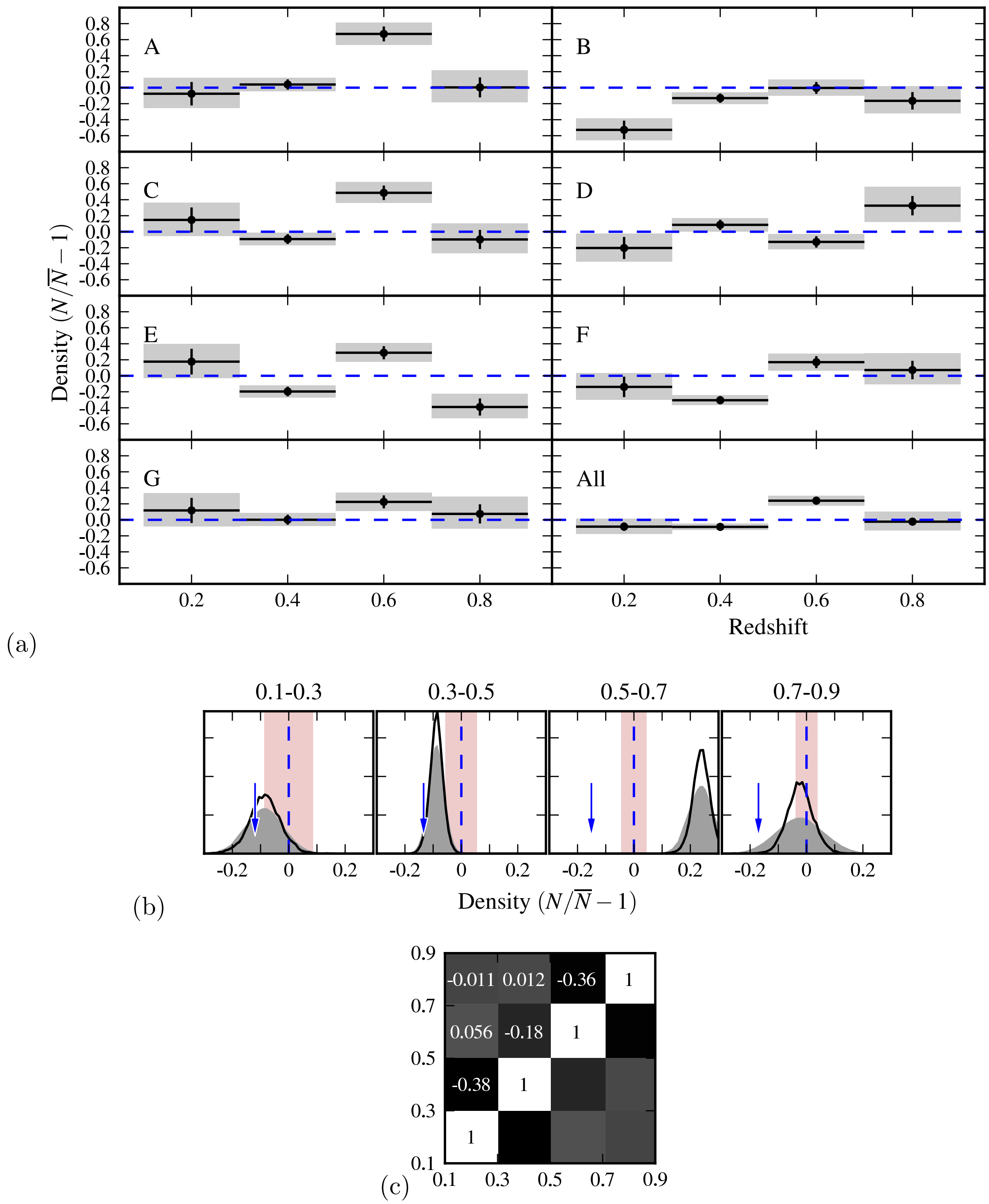

Figure 1: Plotted are the corrected redshift number density distributions reconstructed using our on Monte Carlo method. Panel (a) shows the result from each of the 7 Cold Spot fields (A-G), as well as the mean distribution derived by summing the galaxy counts (All). The error bars give 68% marginalized likelihood ranges. Panel (b) shows the marginalized likelihood distribution of the density for the four redshift bins in the combined measurement. The shaded region is the 1-σ range of cosmic variance and the arrow marks a 200Mpc diameter supervoid with underdensity $\delta = -0.3$. The typical normalized covariance between redshift bins is illustrated in Panel (c). Neighboring bins are anti-correlated by 30%.

1.3 Future Plans

We plan to test how the above non-linear mapping algorithm works under realistic conditions of discrete sampling of the density field, bias and redshift distortion/or photometric redshift.

In addition we have been recently reanalysing our previous algorithms from the point of view of compression. E.g. both the three-point function algorithm and the above mapping can be thought of as a distinct pre-compression phase that precedes the later steps. We plan to systematically look into the question of which algorithms can benefit from a compression phase.

2 Principal Results in 2008

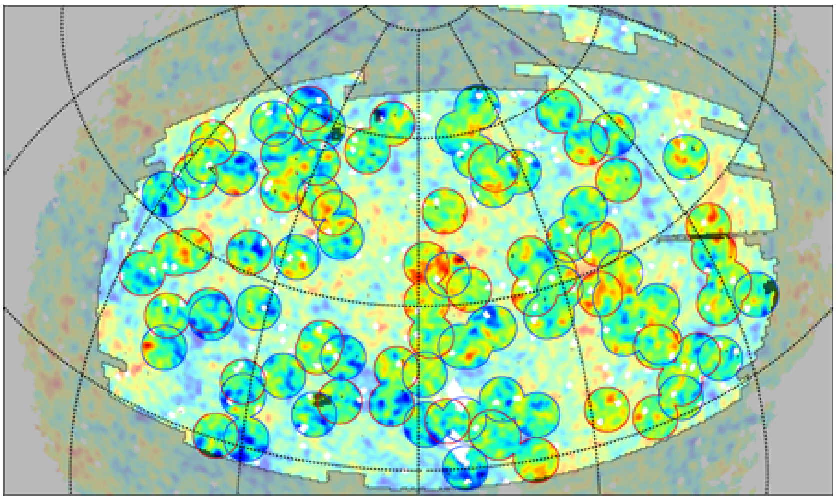

Figure 2: A map of the microwave sky over the SDSS area. The supervoids and superclusters found by our algorithm are highlighted and outlined at a radius of 4°, blue for supervoids and red for superclusters. We used compensated filter to assess the statistical significance of our finding, which approximately corrects for the large-angular-scale temperature variations that are visible across the map. The SDSS DR6 coverage footprint is outlined. Holes in the survey, e.g. due to bright stars, are displayed in black. Additionally, the WMAP Galactic foreground and point source mask is plotted (white holes). The disk of the Milky Way, which extends around the left and right border of the figure, is also masked. The map is in a Lambert azimuthal equal-area projection, centered at right ascension 180°and declination 35°. The longitude and latitude lines are spaced at 30°intervals.

2.1 ISW detection with supervoids and superclusters

The total energy density of the universe today appears to be dominated with Dark Energy (DE), a poorly understood form of energy with negative pressure or anti-gravity. Since DE causes the gravitational potential to decay, photons of the Cosmic Microwave Background (CMB) traveling through a high density region receive a kick. This ultimately results in a hot spot; conversely, low density regions are expected to create cold spots. This phenomenon, the late time integrated Sachse-Wolfe (ISW) effect, has been detected, albeit at low significance, by cross-correlating galaxy surveys with the CMB. This has been done in the past, and SpICE has been one of the major software to carry out the calculation.

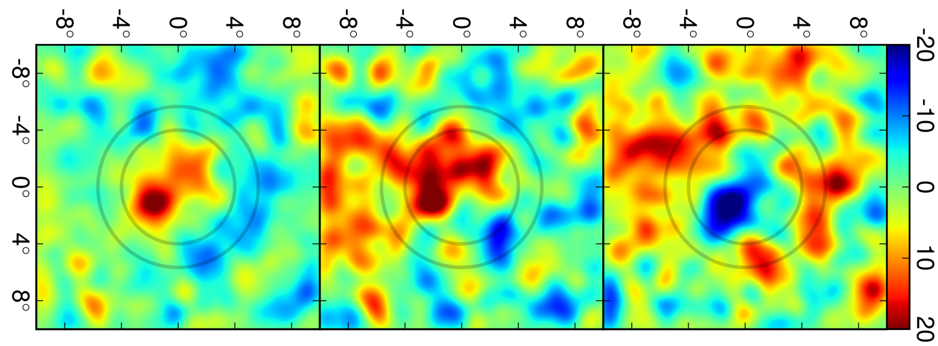

Figure 3: Stacked regions on the CMB corresponding to supervoid and supercluster structures identified in the SDSS LRG catalog. We averaged CMB cut-outs around 50 supervoids (left) and 50 superclusters (middle), and the combined sample (right). The cut-outs are rotated, to align each structure's major axis with the vertical direction. Our statistical analysis uses the raw images, but for this figure we smooth them with a Gaussian kernel with FWHM 1.4°. Hot and cold spots appear in the cluster and void stacks, respectively, with a characteristic radius of 4°, corresponding to spatial scales of 100 h^{-1}Mpc. The inner circle (4° radius) and equal-area outer ring mark the extent of the compensated filter used in our analysis. Given the uncertainty in void and cluster orientations, small-scale features should be interpreted cautiously.

In order detect ISW effect more directly, we developed a pair of algorithms, VOBOZ and ZOBOV to find extreme structures, supervoids and superclusters in galaxy catalogs. We demonstrated the technique in the Sloan Digital Survey photometric Large Red Galaxy survey. We found that supervoids and superclusters, correspond to highly significant cold and hot spots in the Wilkinson Anisotropy Survey CMB map. Stacking areas of the CMB corresponding to 50 galaxy supervoids and 50 superclusters produces an image of these rarest and largest structures. Our findings unambiguously pin this effect to discrete sites with morphology in line with theoretical expectations. The significance is vastly larger (about 4.5σ compared to 2σ for simple cross-correlations with SpICE). We have published our findings in Granett et al. (2008c,b). The final result of our algorithm is an image, which can be simply understood as "possibly imaging Dark Energy";this figure received unusually strong international press coverage, especially in the digital media, such as New Astronomy, National Geographic, etc. Note that we think that the jury is still out whether this image is really due to Dark Energy, but the algorithm, and the underlying statistics, which is what we developed under the AISR program, is solid regardless.

2.2 ISW detection with matched filters

The above algorithm proved that it is possible to find correlations between galaxies and the CMB with a more sophisticated method than simple cross-correlations. The supervoid algorithm has three advantageous properties: uses redshift information, less sensitive to fluctuations in the CMB and employs weighting (albeit a heuristic one).

We tried to formalize the above three properties, and came up with a new maximum likelihood method, based on matched filtering, to replace the above heuristic method with a statistically and computationally more sound one. In our new algorithms, first, we construct a map of the derivative of the gravitational potential traced by galaxies (for first demonstration, again we used the SDSS Luminous Red Galaxies and simulations), the we use the

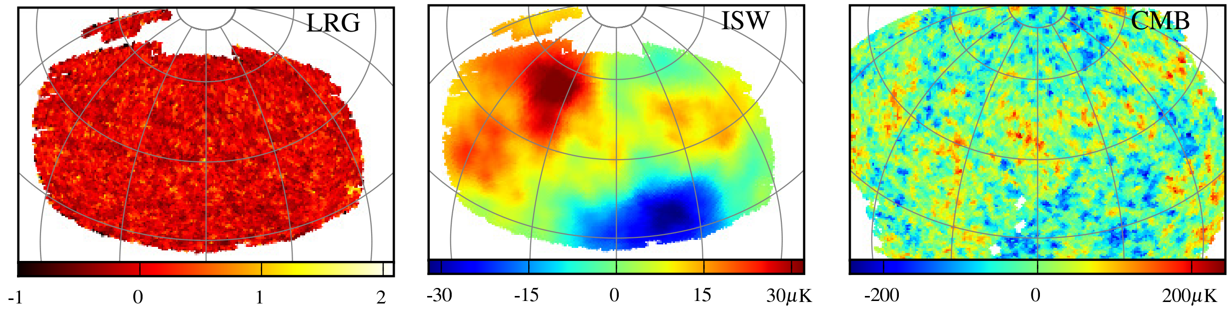

Figure 4: The maps used in our analysis. The left frame is the LRG galaxy overdensity within the SDSS DR6 survey footprint. At center is the reconstructed ISW signal in μK. Right is the foreground-cleaned CMB temperature from the WMAP5 MCMC analysis with the galactic foreground and point source mask applied. The maps are in celestial coordinates with 55 arcminute pixel. The grid spacing is 30° with RA=180° at the center, and RA increasing to the left.

matched filter formalism to asses the significance of the potential present in the CMB. The method produces a 2σ measurement, and only about 10% more efficient than direct cross-correlation with SpICE. Moreover, subtracting the constructed image from the CMB indeed cancels the cross-correlations with SpICE as we demonstrated. While the first application of this technique opens more questions then answer about the supervoid finding, this measurement is consistent with the cross-correlation statistic, strengthening the claim that dark energy is indeed the cause of the correlation. Thus our new algorithm is slightly better statistically, and potentially simplifies the cosmological interpretation. This is on-going work, and needs more software/algorithmic development, but have just submitted the first paper on these results Granett et al. (2008a).

Both of the above projects resulted in further development of our `SpICE` (small improvements and integration with scripts) and `cosmopy` (introduction of ISW related functions) frameworks.

3 Principal Results in 2007

3.1 SpICE and WMAP

We have been continuing to work on our successful software package SpICE and the WMAP specific pipeline based on it. In Chen & Szapudi (2006) we demonstrated the new capabilities by measuring the 2-1 cumulant correlator power spectrum C_l^{21}, a degenerate bispectrum, from the second data release of the Wilkinson Microwave Anisotropy Probe (WMAP). Our high resolution measurements with SpICE span a large configuration space ($\simeq 168 \times 999$) corresponding to the possible cross-correlations of the maps recorded by the different differencing assemblies. We developed a novel method to recover the eigenmodes of the correspondingly large Monte Carlo covariance matrix. We examined its eigenvalue spectrum and used random matrix theory to show that the off diagonal terms are dominated by noise. The care and difficulty we needed to apply for calculating a useful covariance matrix motivated our project on shrinkage estimators (see below). We minimize the χ^2 to obtain constraints for the non-linear coupling parameter $f_{NL} = 22 \pm 52(1\sigma)$.

3.2 Three-point statistics

While the 2-1 cumulant correlator captures a lot of information about the bispectrum, we have been working on calculating the *full* bispectrum based on the three-point function code we developed at the end of the previous AISR. We have extended the numerical methods developed in (Szapudi et al., 2005, also supported by the previous AISR) to turn the problem into a double Hankel-integration. This involved turning these ideas to code by developing a robust numerical Hankel-transformation package using a `python` framework based on the `numpy` and `scipy`. We are on the verge of obtaining the first full bispectrum measurement using this package. Parts of this software package is integrated in our recently developed `cosmopy` framework, a preview (ver 0.3) of which is experimentally available from our group homepage `http://www.ifa.hawaii.edu/cosmowave`.

3.3 Shrinkage Estimators

The experience during the Monte Carlo estimation of the covariance matrix for the 2-1 cumulant correlator power spectrum motivated us to take a hard look at MC covariance matrix estimation: this is a problem which permeates all statistical analysis of the CMB and large scale structure as well. While the focus has been so far on calculating an accurate power spectrum (C_l's), less attention has been focused on the covariance matrix, which is equally important. We have performed full scale Monte Carlo simulations to gauge how errors propagate from the covariance matrix to parameter estimation and we were surprised by the degree to which an otherwise good measurement can yield misleading results due to tiny errors in the covariance matrix.

To remedy the situation, in Pope & Szapudi (2007) we introduced a novel statistical technique, shrinkage estimation, to estimate the power spectrum covariance matrix from a limited number of simulations. Although the idea has been around since the late 50's in the statistics literature, only recent results made it algorithmically feasible to a range of astrophysical problems, in particular covariance matrix estimation. To our knowledge we are the first to apply this technique to astrophysics.

We optimally combined an empirical estimate of the covariance with a model (the target) to minimize the total mean squared error compared to the true underlying covariance. We test our technique on N-body simulations and evaluate its performance by estimating cosmological parameters. Using a simple diagonal target, we show that the shrinkage estimator significantly outperforms both the empirical covariance and the target individually when using a small number of simulations. We find that reducing noise in the covariance estimate is essential for properly estimating the values of cosmological parameters as well as their confidence intervals. We extend our method to the jackknife covariance estimator and again find significant improvement, though simulations give better results. Even for thousands of simulations we still find evidence that our method improves estimation of the covariance matrix. Because our method is simple, requires negligible additional numerical effort, and produces superior results we always advocate shrinkage estimation for the covariance of the power spectrum and other large-scale structure measurements when purely theoretical modeling of the covariance is insufficient. These ideas, although first demonstrated on LSS power spectrum, should be directly applicable to the CMB power spectrum as well.

4 Acknowledged Support

A list of publications acknowledging AISR support is the following:

References

Chen, G., & Szapudi, I. 2006, ApJLett, 647, L87

Granett, B. R., Neyrinck, M. C., & Szapudi, I. 2008a, ArXiv e-prints

Granett, B. R., Neyrinck, M. C., & Szapudi, I. 2008b, ApJLett, 683, L99

Granett, B. R., Neyrinck, M. C., & Szapudi, I. 2008c, ArXiv e-prints, 805

Granett, B. R., Neyrinck, M. C., & Szapudi, I. 2009, ApJ, 701, 414

Neyrinck, M. C., & Szapudi, I. 2007, MNRAS, 375, L51

Neyrinck, M. C., & Szapudi, I. 2008, MNRAS, 384, 1221

Neyrinck, M. C., Szapudi, I., & Szalay, A. S. 2009, ApJLett, 698, L90

Pan, J., Coles, P., & Szapudi, I. 2007, MNRAS, 382, 1460

Pápai, P., & Szapudi, I. 2008, MNRAS, 389, 292

Pope, A. C., & Szapudi, I. 2007, ArXiv e-prints, 711

Pope, A. C., & Szapudi, I. 2008, MNRAS, 389, 766

Szapudi, I., Pan, J., Prunet, S., & Budavári, T. 2005, ApJLett, 631, L1